Saahil Madge

**Independent Work Spring ’15 Checkpoint Report**

This Spring I am investigating how various asset classes and the market in general

respond to specific events. These events can include financial events such as the start of Quantitative Easing, the depegging of the Swiss franc, etc. They may also include political and other non-financial events. The time span analyzed is January 2007 – January 2015.

The first asset class we look at is equities. We use the S&P500 constituents as representative of the US market overall, and so focus on the daily closing stock price of these constituent companies. We set some variable Δ to be a percent change. For each company, we find how many times that company’s price changed by Δ or more in a single day. We then group companies into deciles based on current market cap, and investigate which deciles have been most volatile. Additionally, we can investigate on which dates these price changes occur and plot price as a time series. Events can then be superimposed on this plot, allowing us to see how company-specific or general-market events affected prices.

I have made good progress with the equities investigation. I have collected data from the CRSP database, grouped companies into deciles, and found the dates on which these large price changes occur. Getting the data required familiarizing myself with the financial databases and meeting with Princeton economics data librarians. I was able to group the companies and find price changes by using Python, specifically the *pandas* and *numpy* libraries. Excel was used to survey the data and also to find the market caps, as that was not provided. However, share price and number of shares outstanding were provided, so the market cap can be calculated from these. A major problem I ran into was that the database seems to be giving back extra information. A list of 502 ticker symbols is provided, but data with 546 is received. Some of these extras are due to companies changing their ticker over time, but some are not part of the S&P500 and I cannot figure out why.

The other asset class I have looked at is foreign currencies. I have used the exchange rates of currencies of the top 10 countries by GDP per capita, with all exchange rates against the USD. Other countries that are interesting to look at have also been added, such as China, India, and Canada. Most of these foreign exchange rates have been collected, with only Qatari Rial and Kuwaiti Dinar remaining. We analyze these rates the same way we analyze equities, by setting a Δ threshold and finding all daily rate changes that exceed this threshold. The results indicate that FX rates change far less frequently than the stock prices, which makes sense. Stocks changed more than 15% in a day over 2700 times, whereas currencies changed more than 5% only 27 times.

The first half of the semester has involved obtaining and making sense of all the data. I can now begin to draw conclusions from it. There are three major areas of work. The first involves creating a list of events and plotting them along the time series. The second area is to also look at treasury bills, since they are used as a good representation of interest rates. The third is very open-ended and could lead to some very interesting conclusions. This area involves using computational techniques from machine learning to gain insights from the data. I would like to use *correlation clustering* and other machine learning techniques to investigate models across these different asset classes.